

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-16194

(43) 公開日 平成9年(1997)1月17日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 3/02	3 0 1		G 1 0 L 3/02	3 0 1 D
G 0 6 F 15/18	5 5 0		G 0 6 F 15/18	5 5 0 C
G 1 0 L 7/08			G 1 0 L 7/08	A
9/00			9/00	F
	3 0 1			3 0 1 A

審査請求 未請求 請求項の数11 F D (全 16 頁) 最終頁に続く

(21) 出願番号 特願平7-187966

(22) 出願日 平成7年(1995)6月30日

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72) 発明者 チャン・ジョセフ

東京都品川区北品川6丁目7番35号 ソニー株式会社内

(72) 発明者 西口 正之

東京都品川区北品川6丁目7番35号 ソニー株式会社内

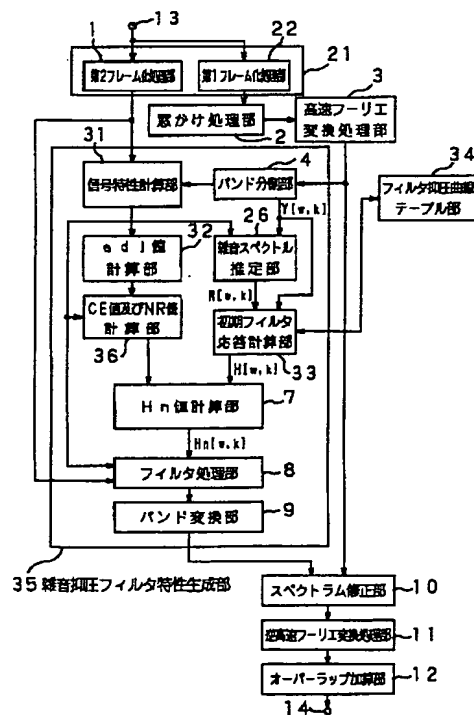
(74) 代理人 弁理士 小池 晃 (外2名)

(54) 【発明の名称】 音声信号の雑音低減方法

(57) 【要約】

【目的】 入力音声信号のピッチ強度の大きいときに所定の帯域の抑圧を抑えることが可能である音声信号の雑音低減方法を提供する。

【構成】 信号特性計算部31にて入力音声信号のピッチ強度を求め、a d j 計算部32にて上記ピッチ強度に応じたa d j 値を求め、C E 値及びN R 値計算部にて上記ピッチ強度に応じたN R 値を求め、H n 値計算部7にて上記N R 値に応じたH n 値を求めて、上記入力音声信号の雑音抑圧量を設定し、スペクトラム修正部10にて上記雑音抑圧量に基づいて上記入力音声信号中の雑音を低減する。



【特許請求の範囲】

【請求項1】 入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給する音声信号の雑音低減方法において、上記所定の周波数帯域における雑音抑圧量を小さくするように周波数特性が制御されることを特徴とする音声信号の雑音低減方法。

【請求項2】 上記音声符号化装置のフィルタは、上記入力される音声信号のピッチ強度に応じて抑圧量に変化するものであり、上記雑音抑圧量を上記入力される音声信号のピッチ強度に応じて変化させることを特徴とする請求項1記載の音声信号の雑音低減方法。

【請求項3】 上記雑音抑圧量は、上記入力される音声信号の高域側の雑音抑圧を小さくするように変化することを特徴とする請求項2記載の音声信号の雑音低減方法。

【請求項4】 上記所定の周波数帯域は、音声信号の低域側であり、上記雑音抑圧量は、上記入力される音声信号の低域側の雑音抑圧を小さくするように変化することを特徴とする請求項1記載の音声信号の雑音低減方法。

【請求項5】 入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給する音声信号の雑音低減方法において、上記入力される音声信号のピッチ強度に応じて、雑音抑圧を行う際の各周波数帯域毎の信号レベルと雑音レベルとの比に対する雑音抑圧特性を変化させることを特徴とする音声信号の雑音低減方法。

【請求項6】 雑音抑圧特性は、雑音抑圧量が上記ピッチ強度が大きいときに小さくなるように制御されることを特徴とする請求項5記載の音声信号の雑音低減方法。

【請求項7】 入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給する音声信号の雑音低減方法において、雑音抑圧の特性を決定する各パラメータをニューラルネットに入力し、上記入力される音声信号の音声区間及び雑音区間の判別が行われることを特徴とする音声信号の雑音低減方法。

【請求項8】 上記ニューラルネットに入力されるパラメータは、少なくとも上記入力される音声信号の自乗平均の平方根及び推定雑音レベルであることを特徴とする請求項7記載の音声信号の雑音低減方法。

【請求項9】 入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給する音声信号の雑音低減方法において、雑音抑圧する際の特性に基づいて処理される最大抑圧量は、dB領域で略線形的に変化することを特徴とする音声信号の雑音低減方法。

【請求項10】 入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号

号を供給する音声信号の雑音低減方法において、上記入力される音声信号のピッチ強度は、信号レベルのピークを選出して得られるピッチ位置の近傍での自己相関を計算することで求められ、

雑音抑圧する際の特性は、上記ピッチ強度に基づいて制御されることを特徴とする音声信号の雑音低減方法。

【請求項11】 入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給する音声信号の雑音低減方法において、

上記入力される音声信号のフレーム化処理は、当該音声信号の特徴を示すパラメータの算出用フレームと、算出されたパラメータを用いてスペクトルの修正を行うフレームとでそれぞれ独立して行われることを特徴とする音声信号の雑音低減方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給する音声信号の雑音低減方法に関する。

【0002】

【従来の技術】携帯用電話機や音声認識等の応用において、收音された音声信号に含まれる環境雑音や背景雑音等の雑音を抑圧し、音声成分を強調することが必要とされている。

【0003】このような音声強調、あるいは雑音低減の技術として、減衰ファクタの調整のために条件付き確率関数を用いる例が、文献「軟判定雑音抑圧フィルタを用いる音声強調」(Speech Enhancement Using a Soft-Decision Noise Suppression Filter, R.J. McAulay, M.L. Malpass, IEEE Trans. Acoust., Speech, Signal Processing, Vol. 28, pp. 137-145, April 1980)や、「移動電話システムにおける周波数領域雑音抑圧研究」(Frequency Domain Noise Suppression Approach in Mobile Telephone Systems, J. Yang, IEEE ICASSP, Vol. II, pp. 363-366, April 1993)等に開示されている。

【0004】

【発明が解決しようとする課題】しかしながら、これらの雑音抑圧技術においては、不適切な固定のSNR(信号対雑音比)に基づく動作をするため、または不適切な抑圧フィルタにより、音色の不自然さや歪んだ音声を生ずることがある。実際の運用の際に、最適の性能を得るために雑音抑圧装置のパラメータの1つであるSNRを調整することは、ユーザにとって望ましいことではない。さらに、従来の音声信号強調技術は、短時間SNRの大きな変動がある音声信号に対して副作用として発生してしまう歪を持たせずに充分雑音を除去することは困難である。

【0005】また、このような音声強調、あるいは雑音低減方法においては、雑音区間検出の技術が用いられ、

入力レベルやパワー等を所定の閾値と比較することにより、雑音区間判別を行っているが、音声にトラッキングすることを防ぐために閾値の時定数を大きくすると、ノイズレベルが変化するとき、特に増加するときに追従できなくなり、誤判別が生じ易くなる。

【0006】ここで、本発明人は、上述した問題を解決するために、特願平6-99869号に記載の明細書及び図面において、音声信号の雑音低減方法を提案している。

【0007】上記音声信号の雑音低減方法は、入力音声信号に基づいて算出された信号レベルと雑音レベルとの比いわゆるSN比及び音声存在確率に基づいて、音声成分を算出するための最尤フィルタを適応的に制御することで雑音抑圧を行う音声信号の雑音低減方法であって、上記音声存在確率の算出に、入力信号のスペクトルから推定雑音スペクトルを減算したものをを用いることを特徴とするものである。

【0008】また、上記音声信号の雑音低減方法によれば、上記最尤フィルタが上記入力音声信号のSN比に応じて最適の抑圧フィルタに調整されるため、上記入力音声信号に対して十分な雑音除去を行うことが可能である。

【0009】ところが、上記音声存在確率を算出するのに、複雑な演算を有すると共に、膨大な演算量が要求されるため、演算の簡略化が望まれる。

【0010】また、例えば音声信号符号化装置に雑音低減装置にて処理を施した音声信号を入力する場合を考えると、上記音声信号符号化装置は予め高域通過フィルタを有していたり、高域側をブーストするフィルタを有しているため、上記雑音低減装置で低域側の抑圧が既になされていると上記音声信号符号化装置においてさらに低域側が抑圧され周波数特性が変化し、聴覚上不自然な音声が出力される虞がある。

【0011】また、従来の雑音低減方法では、雑音低減処理は、入力される音声信号の強さ例えばピッチ強度に基づいてではなく、単に推定される雑音レベルに応じて行われるため、やはり聴覚上不自然な音声が出力される虞がある。

【0012】そこで、上記ピッチ強度を求めるのに、先ずピッチラグを時間波形のピーク値とピーク値との間で求め、このピッチラグで自己相関値を求める方法が知られているが、この場合高速フーリエ変換処理における自己相関関数を用いており、 $(N \log N)$ の項の計算が必要で、さらにNの算出も必要となるため演算が複雑になってしまう。

【0013】そこで、本発明は、上述した実情に鑑みてなされたものであり、入力信号の雑音抑圧を行うのに演算を簡略化することが可能である音声信号の雑音低減方法を提供することを目的とする。

【0014】また、本発明は、特に上記入力音声信号の

ピッチ強度の大きいときに所定の帯域の抑圧を抑えることが可能である音声信号の雑音低減方法を提供することも目的とする。

【0015】

【課題を解決するための手段】本発明は、上述の問題を解決するために、入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給する音声信号の雑音低減方法において、上記所定の周波数帯域における雑音抑圧量を小さくするように周波数特性が制御されるものである。

【0016】また、上記音声符号化装置のフィルタは、上記入力される音声信号のピッチ強度に応じて抑圧量に変化するものであり、上記雑音抑圧量を上記入力される音声信号のピッチ強度に応じて変化させることが挙げられる。

【0017】また、上記所定の周波数帯域は、音声信号の低域側であり、上記雑音抑圧量は、上記入力される音声信号の低域側の雑音抑圧を小さくするように変化することが挙げられる。

【0018】また、本発明は、上述の問題を解決するために、入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給する音声信号の雑音低減方法において、上記入力される音声信号のピッチ強度に応じて、雑音抑圧を行う際の各周波数帯域毎の信号レベルと雑音レベルとの比に対する雑音抑圧特性を変化させるものである。

【0019】また、本発明は、上述の問題を解決するために、入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給する音声信号の雑音低減方法において、雑音抑圧の特性を決定する各パラメータをニューラルネットに入力し、上記入力される音声信号の音声区間及び雑音区間の判別が行われるものである。

【0020】また、本発明は、上述の問題を解決するために、入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給する音声信号の雑音低減方法において、雑音抑圧する際の特性に基づいて処理される最大抑圧量は、dB領域で略線形的に変化するものである。

【0021】また、本発明は、上述の問題を解決するために、入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給する音声信号の雑音低減方法において、上記入力される音声信号のピッチ強度は、信号レベルのピークを選出して得られるピッチ位置の近傍での自己相関を計算することで求められ、雑音抑圧する際の特性は、上記ピッチ強度に基づいて制御されるものである。

【0022】また、本発明は、上述の問題を解決するために、入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給す

る音声信号の雑音低減方法において、上記入力される音声信号のフレーム化处理は、当該音声信号の特徴を示すパラメータの算出用フレームと、算出されたパラメータを用いてスペクトルの修正を行うフレームとでそれぞれ独立して行われるものである。

【0023】

【作用】本発明の音声信号の雑音低減方法によれば、雑音低減に用いるフィルタの特性を制御し、入力される音声信号の所定の周波数帯域における雑音抑圧量を小さくすることで、入力される音声信号の所定の周波数帯域を抑圧するフィルタを有する音声符号化装置に音声信号を供給する。

【0024】また、上記音声符号化装置が、音声信号の低域側を抑圧するフィルタを有している場合、入力される音声信号の低域側の雑音抑圧を小さくするように雑音抑圧量を制御する。

【0025】また、本発明の音声信号の雑音低減方法によれば、入力される音声信号のピッチが検出され、検出されたピッチのピッチ強度が得られる。得られたピッチ強度に応じて、雑音抑圧する際の周波数特性が制御される。

【0026】また、本発明の音声信号の雑音低減方法によれば、雑音抑圧する際の周波数特性を決定する各パラメータをニューラルネットワークに入力すると、入力される音声信号の音声区間及び雑音区間の判別が行われ、この判別は処理を行う回数が増加する程正確になる。

【0027】また、本発明の音声信号の雑音低減方法によれば、入力される音声信号のピッチ強度は、例えば1つの位相内で2つのピークを選出し、各ピークにおける自己相関値とピーク間の相互相関値とを算出し、これら自己相関値及び相互相関値に基づいて得られる。また、上記ピッチ強度に応じて、雑音抑圧する際の周波数特性が制御される。

【0028】また、本発明の音声信号の雑音低減方法によれば、入力される音声信号のフレーム化处理をスペクトルの修正用と、当該音声信号の特徴を示すパラメータの算出用とでそれぞれ独立して行い、例えば上記パラメータの算出用のフレーム化处理を上記スペクトルの修飾用のフレーム化处理よりもサンプル数を多くとるようにする。

【0029】

【実施例】以下、本発明の音声信号の雑音低減方法について、図面を参照しながら説明する。

【0030】ここで、雑音低減装置に適用した本発明の音声信号の雑音低減方法の一例を図1に示す。

【0031】上記雑音低減装置は、雑音抑圧フィルタ特性生成部35において音声信号入力端子13からの入力音声信号の雑音抑圧量を設定し、スペクトラム修正部10にて後述するように、上記雑音抑圧量に基づいて上記入力音声信号中の雑音を低減する。また、音声信号出力

端子14から出力される音声信号は、例えば符号励振線形予測符号化のアルゴリズムにより動作する符号化装置に送られる。

【0032】上記雑音低減装置において、音声信号入力端子13には、音声(Speech)成分と雑音(Noise)成分を含む入力音声信号 $y[t]$ が供給されている。この入力音声信号 $y[t]$ は、例えば標本化周波数がFSのデジタル信号である。り、フレーム化处理部21に送られて、フレーム長がFLサンプルのフレームに分割され、以下各フレーム毎に処理が行われる。

【0033】フレーム化处理部21は、上記入力音声信号のフレーム化处理を、スペクトルの修飾用の処理を行う第1フレーム化处理部22と、当該音声信号の特徴を示すパラメータの算出用の処理を行う第2フレーム化处理部1とでそれぞれ独立して行う部分であり、一方は後述するように雑音抑圧フィルタ特性生成部35に送られ上記入力音声信号の信号特性を示すパラメータを算出するのに用いられ、もう一方は後述するようにスペクトラム修正部10にて上記信号特性を示すパラメータに基づいて得られる雑音抑圧の特性に応じてスペクトルを修正するのに用いられる。

【0034】ここで、第1フレーム化处理部22は、例えば図2のAに示すように、上記入力音声信号を168サンプル単位すなわち上記フレーム長FLが168サンプルのフレームに分割化するいわゆるフレーム化处理を行っており、k番目のフレームを $frame1_k$ として取り出して窓掛け処理部2に出力している。なお、第1フレーム化处理22にて得られる各フレーム $frame1_k$ は160サンプルを周期として取り出され、前後のフレームとは8サンプルのオーバーラップを有している。

【0035】また、第2フレーム化处理部1は、例えば図2のBに示すように、上記入力音声信号を200サンプル単位すなわち上記フレーム長FLが200サンプルとなるようにフレーム化处理しており、k番目のフレームを $frame2_k$ として取り出して信号特性計算部31とフィルタ処理部8とに出力している。なお、第2フレーム化处理部1にて得られる各フレーム $frame2_k$ は、160サンプルを周期として取り出され、1つ前のフレームである $frame2_{k-1}$ とは40サンプルのオーバーラップを有している。

【0036】また、上記周波数やサンプル数の具体例を挙げると、標本化周波数FSを8000Hz、すなわち8kHzとすると、上記第1フレーム化处理部22及び第2フレーム化处理部1ともに、フレーム間隔FIが160サンプルであるため、20ms毎にフレーム化处理が行われることになる。

【0037】図1に戻って、窓かけ処理部2では、次の直交変換である、例えば高速フーリエ変換処理部3での計算に先立って、上記第1フレーム化处理部22より送

られる各フレーム化信号 $y\text{-frame1}_{j,k}$ に対して、窓関数 w_{input} による窓かけ処理が施される。なお、各フレーム毎の信号処理の終段での後述する逆高速フーリエ変換処理のあとには、出力信号に対して窓関数 w_{output} による窓かけ処理が施される。このような各窓関数 w

$$W_{\text{input}}[j] = \left(\frac{1}{2} - \frac{1}{2} \cdot \cos\left(\frac{2 \cdot \pi \cdot j}{FL}\right) \right)^{\frac{1}{4}} \quad 0 \leq j \leq FL \quad \cdots(1)$$

$$W_{\text{output}}[j] = \left(\frac{1}{2} - \frac{1}{2} \cdot \cos\left(\frac{2 \cdot \pi \cdot j}{FL}\right) \right)^{\frac{3}{4}} \quad 0 \leq j \leq FL \quad \cdots(2)$$

【0039】次に、高速フーリエ変換処理部3では、窓関数 w_{input} による窓かけ処理が施されたフレーム化信号 $y\text{-frame1}_{j,k}$ に対して、256ポイントの高速フーリエ変換処理が施され、得られた周波数スペクトル振幅値は、バンド分割部4及びスペクトラム修正部10に出力される。

【0040】また、雑音抑圧フィルタ特性生成部35は、信号特性計算部31、上記 adj 値計算部32、 CE 値及び NRR 値計算部36及び Hn 計算部7を有して成

【0041】上記雑音抑圧フィルタ特性生成部35において、バンド分割部4は、高速フーリエ変換処理部3から出力される入力音声信号を高速フーリエ変換処理して得られる周波数スペクトルの振幅値を、例えば18バンドに分割して、バンドを識別するバンド番号を w とし、各バンドの振幅 $Y[w, k]$ を、信号特性計算部31と雑音スペクトル推定部26と初期フィルタ応答計算部33とに出力する。ここで、バンド分割する際の周波数レンジの一例を次の表に示す。

【0042】

【表1】

*input 及び w_{output} の一例を、次の(1)式及び(2)式にそれぞれ示す。

【0038】

【数1】

バンド番号	周波数範囲
0	0~125Hz
1	125~250Hz
2	250~375Hz
3	375~563Hz
4	563~750Hz
5	750~938Hz
6	938~1125Hz
7	1125~1313Hz
8	1313~1563Hz
9	1563~1813Hz
10	1813~2063Hz
11	2063~2313Hz
12	2313~2563Hz
13	2563~2813Hz
14	2813~3063Hz
15	3063~3375Hz
16	3375~3688Hz
17	3688~4000Hz

【0043】これらの周波数帯域は、人間の聴覚システムが高域ほど知覚分解能が劣化することに基づいている。各帯域の振幅として、対応する周波数レンジ内の最大FFT（高速フーリエ変換処理における周波数帯域の）振幅を用いる。

【0044】また、信号特性計算部31は、第2フレーム化処理部1にて出力される $y\text{-frame2}_{j,k}$ とバンド分割部4にて出力される $Y[w, k]$ とからフレーム毎のRMS値である $RMS[k]$ 、フレーム毎の相対エネルギーである $dBr_{\text{rel}}[k]$ 、フレーム毎の推定雑音レベル値である $MinRMS[k]$ 、フレーム毎の最大RMS値である $MaxRMS[k]$ 、フレーム毎の最大SNR値である $MaxSNR[k]$ を算出する。また、 $y\text{-frame2}_{j,k}$ から後述するピッチを検出し、このピッチ強度を求める。

【0045】まず、上記ピッチの検出とピッチ強度の算出について説明する。

【0046】ピッチの検出において、例えば図3に示すように、入力音声信号 $y\text{-frame}2j,k$ の各フレーム中で1番強いピークがピーク $x[m1]$ として検出され、ピーク $x[m1]$ と同じ位相内で2番目に強いピークがピーク $x[m2]$ として検出される。なお、 $m1$ 及び $m2$ は、各ピークに対応する時間 t の値である。また、ピッチ p の距離がピーク $x[m1]$ 及びピーク $x[m2]$ 間の距離 $|m1-m2|$ として得られる。このピッチ p の最大ピッチ強度 \max_Rxx は、(6) 式に示すように、

(3) 式乃至 (5) 式にて求められるピーク $x[m1]$ とピーク $x[m2]$ との相互相関値 $nrg0$ 、ピーク $x[m1]$ の自己相関値 $nrg1$ 及びピーク $x[m2]$ の自己相関値 $nrg2$ に基づいて得ることができる。

【0047】

【数2】

$$nrg0 = \sum_{\Delta t=-a}^b x[m1+\Delta t] \cdot x[m2+\Delta t] \quad \cdots(3)$$

$$nrg1 = \sum_{\Delta t=-a}^b x[m1+\Delta t] \cdot x[m1+\Delta t] \quad \cdots(4)$$

$$nrg2 = \sum_{\Delta t=-a}^b x[m2+\Delta t] \cdot x[m2+\Delta t] \quad \cdots(5)$$

$$\max_Rxx = \sqrt{\frac{nrg0}{\max(nrg1, nrg2)}} \quad \cdots(6)$$

*

$$dB_{rel}[k] = 10 \log_{10} \left(\frac{E_{decay}[k]}{E[k]} \right) \quad \cdots(8)$$

$$E[k] = \sum_{i=1}^{FL} (y - frame2j, k)^2 \quad \cdots(9)$$

$$E_{decay}[k] = \max \left(E[k], \left(\exp \left(\frac{-FI}{0.65 \cdot FS} \right) \right) \cdot E_{decay}[k-1] \right) \quad \cdots(10)$$

【0053】また、上記(10)式においては、減衰時間(ディケイタイム)を0.65秒とした例を示している。

【0054】このような、エネルギー $E[k]$ 及び減衰エネルギー $E_{decay}[k]$ の具体例を、図4に示す。

【0055】また、第 k フレーム $frame2k$ の最大RMS値 $\max RMS[k]$ は、後述する各フレームの推定雑音レベル値と最大SN比とを見積もるのに必要な値であって、以下の(11)式にて算出される。(11)式で、 θ は減衰定数(decay constant)であり、例えば3.2秒で最大RMS値が $1/e$ だけ減衰するような値、すなわち $\theta = 0.993769$ が用いられる。

【0056】

【数5】

$$\max RMS[k] = \max(4000, RMS[k], \theta \cdot \max RMS[k-1] + (1-\theta) \cdot RMS[k]) \quad \cdots(11)$$

【0057】第 k フレーム $frame2k$ の推定雑音レベル値 $\min RMS[k]$ は、バックグラウンドノイズ、ある

*【0048】続いて、上記各値の算出方法を説明する。

【0049】RMS $[k]$ は、第 k フレーム $frame2k$ の RMS 値であって、例えば次式で計算される。

【0050】

【数3】

$$RMS[k] = \sqrt{\frac{1}{FL} \cdot \sum_{j=0}^{FL-1} (y - frame2j, k)^2} \quad \cdots(7)$$

【0051】第 k フレーム $frame2k$ の相対エネルギー $dB_{rel}[k]$ は、前フレーム $frame2k-1$ からの減衰エネルギーに関連する第 k フレームの相対エネルギーを示すものであって、この dB 表示の相対エネルギー $dB_{rel}[k]$ は、例えば次の(8)式により計算され、この(8)式中のエネルギー $E[k]$ 及び減衰エネルギー値 $E_{decay}[k]$ は、それぞれ次の(9)式及び(10)式により求められる。

【0052】

【数4】

20

いは背景雑音のレベルを評価するのに好適な最小の RMS 値であって、現時点から前に5個の局所極小値(local minimum)、すなわち(12)式を満たす値の内での最小となる値である。

【0058】

【数6】

$$\begin{aligned} (RMS[k] < 0.6 \cdot \max RMS[k]) & \text{ かつ} \\ RMS[k] < 4000 & \text{ かつ} \\ RMS[k] < RMS[k+1] & \text{ かつ} \\ RMS[k] < RMS[k-1] & \text{ かつ} \\ RMS[k] < RMS[k-2] & \text{ 又は} \\ (RMS[k] < \min RMS) & \cdots(12) \end{aligned}$$

【0059】この推定雑音レベル値 $\min RMS[k]$ は、音声(Speech)無しの背景雑音、いわゆるバックグラウンドノイズのときに上昇してゆくように設定されている。ノイズレベルが高いときの上昇レートは指数関数的であるが、低いノイズレベルのときには、より大きな上昇を得るために固定の上昇レートが用いられる。

11

【0060】これらのRMS値RMS[k]、推定雑音レベル値MinRMS[k]及び最大RMS値MaxRMS[k]の具体例を、図5に示す。

【0061】第kフレームframe2kの最大SN比MaxSNR[k]は、MaxRMS[k]とMinRMS[k]とに基づいて、以下の(13)式により推定される値である。

【0062】

【数7】

$$NR_level[k] = \begin{cases} \left(\frac{1}{2} + \frac{1}{2} \cos \left(\pi \cdot \frac{MaxSNR[k] - 30}{20} \right) \right) \times \\ \left(1 - 0.002(MaxSNR[k] - 30)^2 \right) & 30 < MaxSNR[k] \leq 50 \\ 0.0 & MaxSNR[k] > 50 \\ 1.0 & \text{上記以外} \end{cases} \quad \cdots(14)$$

【0065】次に、雑音スペクトル推定部26は、RMS[k]、dBrel[k]、NR_level[k]、MinRMS[k]及びMaxSNR[k]に基づいて、音声(speech)を背景雑音(background noise)から区別する。すなわち、次の条件が正しいとき、第kフレーム中の信号は背景雑音として分類される。こうして分類され※

((RMS[k] < NoiseRMS_{thres}[k])) 又は

(dBrel[k] > dB_{thres}[k]) かつ

(RMS[k] < RMS[k-1] + 200) …(15)

ここで NoiseRMS_{thres}[k] = 1.05 + 0.45 · NR_level[k] ×

MinRMS[k]

$$dB_{thres_rel}[k] = \max(MaxSNR[k] - 4.0, 0.9 \cdot MaxSNR[k])$$

【0067】ここで図6は、上記(15)式中のdB表示の相対エネルギーdBrel[k]と、最大SN比MaxSNR[k]と、雑音判別の閾値の1つであるdB_{thres}rel[k]との具体例を示している。

【0068】また図7は、上記(14)式中のMaxSNR[k]の関数としてのNR_level[k]を示している。

【0069】第kフレームが背景雑音、あるいはノイズとして分類される場合、上記ノイズスペクトルの時間平均推定値N[w, k]は、現在フレームの信号の入力信号スペクトルの振幅Y[w, k]によって、次の(16)式のように更新される。なお、wは上記バンド分割のバンド番号を示すものである。

【0070】

【数10】

$$adj[w, k] = \min(adj1[k], adj2[k]) - adj3[w, k] \quad \cdots(17)$$

12

$$* \quad MaxSNR[k] = 20 \cdot \log_{10} \left(\frac{MaxRMS[k]}{MinRMS[k]} \right) - 1 \quad \cdots(13)$$

【0063】また、この最大SN比值MaxSNRからは、相対ノイズレベルを示す0から1までの範囲の正規化パラメータNR_level[k]が算出される。このNR_level[k]には、以下の関数を用いられる。

【0064】

【数8】

※た背景雑音を示す振幅値は、ノイズスペクトルの時間平均推定値N[w, k]として算出され、初期フィルタ応答計算部33に出力される。

【0066】

【数9】

$$N[w, k] = \alpha \cdot \max(N[w, k-1], Y[w, k]) + (1 - \alpha) \cdot \min(N[w, k-1], Y[w, k]) \quad \cdots(16)$$

$$\alpha = \exp \left(\frac{-FI}{0.5 \cdot FS} \right)$$

【0071】ここで、第kフレームが音声(speech)として分類された場合、N[w, k]はN[w, k-1]の値をそのまま用いる。

【0072】次に、adj値計算部32は、RMS[k]、MinRMS[k]及びMaxRMS[k]に基づいて、後述するadj1[k]、adj2[k]及びadj3[w, k]を用いて、(17)式によりadj[w, k]を算出し、CE値及びNR値計算部36に出力する。

【0073】

【数11】

13

14

【0074】ここで、(17)式中の $adj1[k]$ は、全ての帯域において、高いSN比における後述するフィルタ処理による雑音抑圧動作を抑える効果を有する*

*値であり、以下の(18)式にて定義される。

【0075】

【数12】

$$adj1[k] = \begin{cases} 1 & MaxSNR[k] < 29 \\ 1 - \frac{MaxSNR[k] - 29}{14} & 29 \leq MaxSNR[k] < 43 \\ 0 & \text{上記以外} \end{cases} \quad \dots(18)$$

【0076】また、(17)式中の $adj2[k]$ は、

※る。

非常に低いノイズレベルや非常に高いノイズレベルに対

10 【0077】

して、上記フィルタ処理による雑音抑圧レートを抑える

【数13】

効果を有する値であり、以下の(19)式にて定義され※

$$adj2[k] = \begin{cases} 0 & MinRMS[k] < 20 \\ \frac{MinRMS[k] - 20}{40} & 20 \leq MinRMS[k] < 60 \\ 1 & 60 \leq MinRMS[k] < 1000 \\ 1 - \frac{(MinRMS[k] - 1000)}{1000} & 1000 \leq MinRMS[k] < 1800 \\ 0.2 & MinRMS[k] \geq 1800 \end{cases} \quad \dots(19)$$

【0078】また、(17)式中の $adj3[w, k]$

は、図3で示したような入力音声信号のピッチ p の強度、特に最大ピッチ強度 max_Rxx が大きいに、低域側と高域側とにおける雑音抑圧量を低減するように制御する値であり、例えばピッチ強度が所定値より大きくかつ入力音声信号レベルが雑音レベルより大きい場合は、図8のAに示すような低域側で所定値をとり高域側で周波数 w に対して線形的に変化するとともに他の周波★

★数帯域では0である値であり、また、その他の場合は、図8のBに示すように低域側で所定値をとるとともに他の周波数帯域では0である値である。

【0079】この $adj3[w, k]$ の定義の一例を、(20)式に示す。

【0080】

【数14】

$$\frac{max_Rxx[t]}{max_Rxx[0]} > 0.55 \text{ かつ } RMS[k] > 0.8 \cdot MinRMS[k] + 0.2 \cdot MaxRMS[k]$$

$$adj3[w, k] = \begin{cases} 0.2 & w < 200 \text{ Hz} \\ 0 & 200 \leq w < 2375 \text{ Hz} \\ \frac{0.059415(w - 2375)}{4000 - 2375} & w \geq 2375 \text{ Hz} \end{cases}$$

上記以外

$$adj3[w, k] = \begin{cases} 0.2 & w < 200 \text{ Hz} \\ 0 & w \geq 200 \text{ Hz} \end{cases} \quad \dots(20)$$

【0081】この(20)式において、最大ピッチ強度 $max_Rxx[t]$ は、最初の最大ピッチ強度 $max_Rxx[0]$ で規格化して用いており、また、入力音声レベルと雑音レベルとの比較は、 $MinRMS[k]$ 及び $MaxRMS[k]$ を用いて得られる値を用いて行われる。

【0082】また、CE値及びNR値計算部36は、フ

ィルタ特性を制御するNR値を求めてHn値計算部7に出力する。

【0083】ここで、例えば上記NR値であるNR $[w, k]$ は、以下の(21)式で定義される。

【0084】

【数15】

$$NR[w, k] = (1.0 - CE[k]) \cdot NR'[w, k] \quad \cdots (21)$$

$$NR'[w, k] = \begin{cases} adj[w, k] & NR[w, k-1] - \delta_{NR} \\ & < adj[w, k] < NR[w, k-1] + \delta_{NR} \\ NR[w, k-1] - \delta_{NR} & NR[w, k-1] - \delta_{NR} \geq adj[w, k] \\ NR[w, k-1] + \delta_{NR} & NR[w, k-1] + \delta_{NR} \leq adj[w, k] \end{cases}$$

$$\delta_{NR} = 0.004 \quad \cdots (22)$$

【0085】(21)式の $NR'[w, k]$ は、上記 adj 値計算部32から送られる $adj[w, k]$ を用いて(22)式にて求められる値である。

【0086】また、上記 CE 値及び NR 値計算部36は、(21)式で用いる $CE[k]$ の算出も行っている。この $CE[k]$ は、上記入力信号スペクトルの振幅 $Y[w, k]$ 中に含まれる子音成分をフレーム毎に検出した子音検出結果を表す値である。この子音検出処理の具体例について説明する。

【0087】上記ピッチ強度が所定値より大きくかつ入力音声信号レベルが雑音レベルより大きい場合、すなわち(20)式の最初に示されている条件が満たされる場合、 $CE[k]$ は、例えば0.5をとる。また、上記条件を満たさない場合は、以下の方法で決定される $CE[k]$ を用いる。

【0088】まず、上記 $Y[w, k]$ 中の連続するサンプル間で符号が、例えば正から負、あるいは負から正というように逆転する箇所、または、逆の符号を有するサンプル間で0という値を有するサンプルが存在する箇所がゼロクロスとして検出される。このゼロクロスの数が、フレーム毎に検出され、この値がゼロクロス数 $ZC[k]$ として以下の処理に用いられる。

【0089】次に、トーン、すなわち上記 $Y[w, k]$ の周波数成分の分布を表す値、例えば、図9に示すように、高域における上記入力信号スペクトルの平均レベル t' と低域における上記入力信号スペクトルの平均レベ

*ル b' との比 t'/b' (= $tone[k]$)が検出される。この値 t' 及び値 b' は、以下の(23)式にて定義される誤差関数 $Err(fc, b, t)$ が最小値をとるような値 t 及び値 b である。(23)式において、 NB はバンド数を表し、 $Y_{max}[w, k]$ はバンド w における $Y[w, k]$ の最大値を表し、 fc は高域と低域とを分離する点を表す。また、図9において、周波数 fc において、低域側の $Y[w, k]$ の平均値を値 b として、また、高域側の $Y[w, k]$ の平均値を値 t として

【0090】

【数16】

$$\min_{\substack{fc=2 \dots NB-3 \\ b, te R}} Err(fc, b, t) = \sum_{w=0}^k (Y_{max}[w, k] - b)^2 + \sum_{w=fc+1}^{NB-1} (Y_{max}[w, k] - t)^2 \quad \cdots (23)$$

【0091】さらに、RMS値及びゼロクロス数に基づいて、有声音声が検出されたフレームの近傍のフレーム、すなわち音声近接フレームが検出され、このフレーム数として音節近接フレーム数 $spch_prox[k]$ が、以下の(24)式に基づいて得られ、出力される。

【0092】

【数17】

$$spch_prox[k] = \begin{cases} 0 & (RMS_i > 1250) \quad (ZC_i < 70) \\ & \text{ただし } i = k-4, \dots, k \\ spch_prox[k-1] & \text{上記以外} \end{cases} \quad \cdots (24)$$

【0093】そして、上記ゼロクロス数、上記音声近接フレーム数、上記トーン及びRMS値に基づいて、各フレームの $Y[w, k]$ 中の子音成分の検出が行われる。この子音検出結果として $CE[k]$ が、以下の(25)

$$CE[k] = \begin{cases} E & (tone[k] > 0.6) \text{ かつ } (C1, C2, \text{及び } C3 \text{ が真}) \\ & \text{かつ } (C4.1, C4.2, \dots, \text{又は } C4.7 \text{ が真}) \\ \max\{0, CE[k-1] - 0.05\} & \text{上記以外} \end{cases} \quad \cdots (25)$$

【0095】また、各シンボル $C1$ 、 $C2$ 、 $C3$ 、 $C4.1$ 乃至 $C4.7$ は、以下の表にて定義される。

【0096】

【表2】

式に基づいて得られる。

【0094】

【数18】

シンボル	定義式
C1	$\text{RMS}[k] > \text{CDS0} \cdot \text{MinRMS}[k]$
C2	$\text{ZC}[k] > \text{Z low}$
C3	$\text{spch_prox}[k] < T$
C4.1	$\text{RMS}[k] > \text{CDS1} \cdot \text{RMS}[k-1]$
C4.2	$\text{RMS}[k] > \text{CDS1} \cdot \text{RMS}[k-2]$
C4.3	$\text{RMS}[k] > \text{CDS1} \cdot \text{RMS}[k-3]$
C4.4	$\text{ZC}[k] > \text{Z high}$
C4.5	$\text{tone}[k] > \text{CDS2} \cdot \text{tone}[k-1]$
C4.6	$\text{tone}[k] > \text{CDS2} \cdot \text{tone}[k-2]$
C4.7	$\text{tone}[k] > \text{CDS2} \cdot \text{tone}[k-3]$

【0097】上記表2において、CDS0、CDS1、CDS2、T、Z low及びZ highの各値は、子音検出の感度を決定する定数であり、例えばCDS0=CDS1=CDS2=1.41、T=20、Z low=20、Z high=75の値をとる。また、(25)式のEは、0から1までの値をとるもので、0に近いほど通常の子音抑圧量に近くなるように後述するフィルタ応答が調整され、また、1に近いほど子音抑圧量が最低量となるように上記フィルタ応答が調整される値であり、例えば0.7が用いられる。

【0098】また、上記表2によれば、あるフレームにおいて、シンボルC1が成立することは上記フレームの信号レベルが最低ノイズレベルより大きいことを示し、シンボルC2が成立することは上記フレームのゼロクロス数が所定のゼロクロス数Z low、本実施例では20より大きいことを示し、また、シンボルC3が成立することは上記フレームが有声音声が検出されたフレームよりTフレーム以内、本実施例では20フレーム以内であることを示している。

【0099】また、シンボルC4.1が成立することは上記フレームにおいて信号レベルが変化することを示し、シンボルC4.2が成立することは上記フレームが音声信号が変化して1フレーム後であって信号レベルが変化するフレームであることを示し、また、シンボルC4.3が成立することは上記フレームが音声信号が変化して2フレーム後であって信号レベルが変化するフレームであることを示している。また、シンボルC4.4が成立することは、上記フレームにおいて、ゼロクロス数が所定のゼロクロス数Z high、本実施例では75より大きいことを示している。また、シンボルC4.5が

成立することは上記フレームにおいてトーン値が変化す

ることを示し、シンボルC4.6が成立することは上記

フレームにおいてトーン値が変化して1フレーム後であってト

ーン値が変化することを示し、シンボル

C4.7が成立することは上記フレームが音声信号が変

化してトーン値が変化するフレー

ムである。

【0100】また、(25)式によれば、このフレームが子音成分を含んでいることの条件は、上述のシンボルC1乃至C3の条件を満たすこと、tone[k]が0.6より大きいこと及び上述のC4.1乃至C4.7の条件の内の少なくとも1つが満たされることである。

【0101】また、初期フィルタ応答計算部33は、雑音スペクトル推定部26から出力される雑音時間平均値N[w, k]と、バンド分割部4から出力されるY[w, k]とをフィルタ抑圧曲線テーブル部34に送り、フィルタ抑圧曲線テーブル部34に収納されるY[w, k]とN[w, k]とに応じたH[w, k]の値を探し出し、このH[w, k]をHn値計算部7に出力する。なお、フィルタ抑圧曲線テーブル部34は、H[w, k]に関する表が格納されている。

【0102】Hn値計算部7は、バンド分割された入力信号スペクトルの振幅Y[w, k]と、ノイズスペクトルの時間平均推定値N[w, k]と、上記NR[w, k]とから、上記バンド分割された入力信号スペクトルの振幅Y[w, k]から雑音成分を低減するためのプレフィルタである。ここでは、Y[w, k]がN[w, k]に応じてHn[w, k]に変換され、このフィルタ応答Hn[w, k]が出力される。なお、このHn[w, k]値は、以下の(26)式に基づいて算出される。

【0103】

【数19】

$$Hn[w, k] = \exp\{NR[w, k] \cdot \ln(H[w] [S/N = r])\} \dots (26)$$

$$20 \cdot \log_{10}(H[w, k]) = NR[w, k] \cdot K \dots (27)$$

但しKは定数

【0104】また、上記(26)式中の値H[w] [S/N=r]は、SN比をある値rに固定したとき最適なノイズ抑圧フィルタ特性に当たり、この値は、Y[w, k]/N[w, k]の値に応じてテーブル化されており、上記フィルタ抑圧曲線テーブル部34に格納されている。なお、上記H[w] [S/N=r]は、dB領域で直線的に変化する値である。

【0105】さらに、上記(26)式を(27)式のように変形すると、最大抑圧量の関数である左辺は、NR[w, k]と直線関係にあることが示され、両者の関係を図10のように示すことができる。

【0106】また、フィルタ処理部8では、上記Hn[w, k]値が周波数軸方向と時間軸方向とについて円滑化するフィルタ処理を行い、得られる信号として円滑化信号Ht_smooth[w, k]が出力される。上記周波数軸方向へのフィルタ処理は、Hn[w, k]の有効インパルス応答長を短くする効果がある。これにより周波数領域での乗算によるフィルタの実現に起因する環状畳み込みによるエリアシングの発生を未然に防いでいる。また、上記時間軸方向へのフィルタ処理は、突発的な雑音

を抑えるフィルタの変化の速さを制限する効果がある。

【0107】 先ず、上記周波数軸方向へのフィルタ処理についての説明を行う。上記各バンドの $H_n[w, k]$ に、メディアン（中央値）フィルタ処理が施される。次の（28）式及び（29）式にて、この方法を示す。

【0108】

【数20】

$$\text{step1: } H1[w, k] = \max\{\text{median}(Hn[w-1, k], Hn[w, k], H[w+1, k]), Hn[w, k]\}$$

但し、 $(w-1)$ 又は $(w+1)$ が存在しない場合、 $H1[w, k] = Hn[w, k]$... (28)

$$\text{step2: } H2[w, k] = \min\{\text{median}(H1[w-1, k], H1[w, k], H1[w+1, k]), H1[w, k]\}$$

但し、 $(w-1)$ 又は $(w+1)$ が存在しない場合、 $H2[w, k] = H1[w, k]$... (29)

*

【数21】

$$H_{\text{speech}}[w, k] = 0.7 \cdot H2[w, k] + 0.3 \cdot H2[w, k-1] \quad \dots (30)$$

$$H_{\text{noise}}[w, k] = 0.7 \cdot \text{Min}_H + 0.3 \cdot \text{Max}_H \quad \dots (31)$$

$$\text{但し、} \text{Min}_H = \min(H2[w, k], H2[w, k-1])$$

$$\text{Max}_H = \max(H2[w, k], H2[w, k-1])$$

【0112】 また、背景雑音の信号に対しては、次の（31）式に示すような時間軸での円滑化、あるいはスムージングを行う。

【0113】 また、過度的状態の信号に対しては、この時間軸でのスムージングを行われない。

【0114】 以上のスムージング処理が行われた信号を用いて、（32）式により円滑化出力信号 $H_{t_smooth}[w, k]$ を得る。

【0115】

【数22】

* 【0109】 （28）式の第1段階（Step1）において、 $H1[w, k]$ は、単一の、あるいは孤立した0のバンドを無くした $Hn[w, k]$ であり、（29）式の第2段階（Step2）において、 $H2[w, k]$ は、単一の、あるいは孤立した突出したバンドを無くした $H1[w, k]$ である。このようにして、上記 $Hn[w, k]$ は、 $H2[w, k]$ に変換される。

【0110】 次に、上記時間軸方向へのフィルタ処理についての説明を行う。この時間軸方向へのフィルタ処理を施す際において、入力信号には、音声（speech）、バックグラウンドノイズ、そして音声（speech）の立ち上がり部分である過度的状態の3種あることを考慮に入れる。音声の信号 $H_{\text{speech}}[w, k]$ に対しては、次の（30）式に示すように、時間軸での円滑化、あるいはスムージングを行う。

【0111】

【数21】

$$H_{t_smooth}[w, k] = (1 - \alpha_{tr}) \cdot \{\alpha_{sp} \cdot H_{\text{speech}}[w, k] + (1 - \alpha_{sp}) \cdot H_{\text{noise}}[w, k]\} + \alpha_{tr} \cdot H2[w, k] \quad \dots (32)$$

$$\alpha_{sp} = \begin{cases} 1.0 & \text{SNR}_{\text{inst}} > 4.0 \\ (\text{SNR}_{\text{inst}} - 1) \cdot \frac{1}{3} & 1.0 < \text{SNR}_{\text{inst}} < 4.0 \\ 0 & \text{上記以外} \end{cases} \quad \dots (33)$$

$$\text{ここで } \text{SNR}_{\text{inst}} = \frac{\text{RMS}[k]}{\text{MinRMS}[k]}$$

$$\alpha_{tr} = \begin{cases} 1.0 & \delta_{rms} > 3.5 \\ (\delta_{rms} - 2) \cdot \frac{2}{3} & 2.0 < \delta_{rms} < 3.5 \\ 0 & \text{上記以外} \end{cases} \quad \dots (34)$$

$$\text{ここで } \delta_{rms} = \frac{\text{RMS}_{\text{local}}[k]}{\text{RMS}_{\text{local}}[k-1]}$$

$$\text{RMS}_{\text{local}}[k] = \sqrt{\frac{1}{FI} \cdot \sum_{j=FI/2}^{FI-FI/2} (y - \text{frame2j}, k)^2}$$

【0116】 ここで、（32）式中の α_{sp} は次の（33）式から、 α_{tr} は次の（34）式からそれぞれ求められる。

【0117】 続いて、バンド変換部9では、フィルタ処理部8からの、例えば18バンド分の円滑化信号 $H_{t_smooth}[w, k]$ が、例えば128バンド分の信号 $H_{128}[w, k]$ に、補間処理により拡張変換され、この

変換された信号 $H_{128} [w, k]$ が出力される。この変換は、例えば2段階で行っており、18バンドから64バンドへの拡張はゼロ次ホールドにより、64バンドから128バンドへの拡張はローパスフィルタ型の補間処理により、それぞれ行っている。

【0118】次に、スペクトラム修正部10では、高速フーリエ変換処理部3で得られたフレーム化信号 $y_{frame,j,k}$ の高速フーリエ変換処理にて得られるFFT係数の実部と虚部とに各々上記信号 $H_{128} [w, k]$ を乗じてスペクトラム修正、すなわち雑音成分を低減する処理が行われ、得られた信号が出力される。この結果、スペクトルの振幅は修正されるが位相は変形を受けない。

【0119】次に逆高速フーリエ変換処理部11では、スペクトラム修正部10にて得られた信号を用いて、逆高速フーリエ変換処理が行われ、得られたIFFT信号が出力される。

【0120】次に、オーバーラップ加算部12では、各フレーム毎のIFFT信号のフレーム境界部分についての重ね合わせが行われ、得られた出力音声信号が音声信号出力端子14より出力される。

【0121】さらに、この出力を例えば符号励振線形予測符号化のアルゴリズムに用いた場合を考える。

【0122】ここで、上記符号励振線形予測符号化のアルゴリズムによる符号化装置を図11に、また、復号化装置を図12にそれぞれ示す。

【0123】上記符号化装置は、図11に示すように、入力端子61から入力音声信号が入力され、線形予測符号分析またはLPC (linear predictive coding) 分析部62及び減算器64に送られる。

【0124】LPC分析部62は、上記入力音声信号の線形予測を行いこの予測フィルタ係数を合成フィルタ63に出力する。合成フィルタ63は、2つのコードブックの1つである固定コードブック67からのコードワードに乗算器81の利得を掛けたデータと、ダイナミックコードブック68からのコードワードに乗算器82の利得とを掛けたデータとを加算器69にて加算された出力が、上記LPC分析部62から送られる予測フィルタ係数を持つLPC合成フィルタに入力され、その合成出力が、減算器64に出力する。

【0125】また、減算器64は、上記入力音声信号と合成フィルタ63からの合成出力との差を取り出し聴覚重み付けフィルタ65に出力する。聴覚重み付けフィルタ65は、周波数帯域毎に入力音声信号のスペクトルに応じた重みを付けて、誤差検出部66に出力する。誤差検出部66は、聴覚重み付けフィルタ65からの出力の重み付き誤差のエネルギーを算出して、固定コードブック67及びダイナミックコードブック68のコードブックサーチにおいて、この重み付き誤差エネルギーが最小になるような各コードブックのコードワードが取り出され

る。

【0126】上記符号化装置からは上記固定コードブック67のコードワードのインデックス、ダイナミックコードブック68のコードワードのインデックス、各乗算器に対応する利得のインデックス、LPC分析部62からフィルタ係数を生成する各パラメータの量子化インデックス等が復号化装置に伝送され、復号化装置にて各インデックスを用いた復号化処理がなされる。

【0127】そこで、上記復号化装置は、図12に示すように、固定コードブック71は上記符号化装置から送られる上記固定コードブック67のコードワードのインデックスに基づいて、ダイナミックコードブック72は上記ダイナミックコードブック68のコードワードのインデックスに基づいて、各コードワードを固定コードブック71あるいはダイナミックコードブック72より取り出す。また、乗算器83、84は、各対応する利得インデックスに基づいて動作し、合成フィルタ74は、上記量子化インデックス等の各パラメータが送られ、これらパラメータを用いて、2つのコードブックからのコードワードに利得が乗算されたデータを励起信号と合成した合成出力をポストフィルタ75に出力する。ポストフィルタ75では、いわゆるフォルマント強調を行い、信号の谷間と山とをより明確にする操作が行われる。フォルマント強調がなされた音声信号が出力端子76より出力される。

【0128】また、ここで、例えば聴覚上より好ましい音声信号を得るために、上記アルゴリズムでは、符号化する音声信号の低域側を抑圧したり、高域側をブーストするフィルタ処理が含まれており、上記復号化装置からのデコーダ出力信号は、低域側が抑圧された音声信号となっている。

【0129】従って、上述のように本発明の音声信号の雑音低減方法では、上記フィルタ処理に応じて、adj値計算部32におけるadj3[w, k]の値を、特にピッチが大きい音声信号に対しては低域側で所定値を有し、さらに高域側で周波数に対して線形関係を有するように見積もることで、結果的に低域側の音声信号の抑圧が抑えられるため、上記アルゴリズムの処理によるフォルマント強調がなされた音声信号の低域側は、過度な抑圧を受ける虞がない、すなわち符号化処理等により本来の周波数特性の変化を減少させることができる。

【0130】なお、ここでは、上記雑音低減装置が、低域側の音声信号を抑圧したり、高域側ブーストするようなフィルタ処理を行う音声符号化装置に出力する例を挙げたが、雑音抑圧する際の高域側の音声信号の抑圧を抑えるようにadj3[w, k]を設定することで、例えば高域側の音声信号を抑圧するような音声符号化装置に出力する場合にも用いることが可能である。

【0131】また、CE値及びNR値計算部36において、CE値をピッチの大きさに応じて計算方法を変え

て、このCE値に応じてNR値を決定するため、ピッチ強度に応じたNR値を得ることが可能であり入力される音声信号に即したNR値により雑音抑圧を行うことが可能になるため、スペクトル量子化誤差を減少させることができる。

【0132】また、Hn値計算部7において、Hn [w, k] を入力されるNR [w, k] に対して、dB領域で略直線的に変化させることで、Hn値の変化に対するNR値の寄与は常に連続しており、急激にNR値が変化してもHn値の変化はこれに対応する。

【0133】また、信号特性計算部31にて最大ピッチ強度を算出するのに、例えば高速フーリエ変換処理において用いる $(N+1)\log N$ といった自己相関関数による複雑な計算を行う必要がなくなり、例えば200サンプルの処理を行った場合、上記自己相関関数では50000回の処理が必要であったのに対して、本発明では30000回の処理で済むため、演算処理の速度を上げることができる。

【0134】また、図2のAに示したように、第1フレーム化処理部22のフレーム化処理ではフレーム長FLが168サンプル、各フレームは前後のフレームと8サンプルずつオーバーラップ部分を有するようにサンプリングされ、また、図2のBに示したように、第2フレーム化処理部1のフレーム化処理をフレーム長FLが200サンプル、各フレームは1つ手前のフレームとは40サンプル、1つ先のフレームとは8サンプルのオーバーラップ部分を有するようにサンプリングさせて、かつ、第1フレーム化処理部22と第2フレーム化処理部1とで各フレームの開始位置を同じにして、上記第2フレーム化処理部1の方が上記第1フレーム化処理部22よりも32サンプル分だけ後ろにずらすことで、第1フレーム化処理部22及び第2フレーム化処理部1間で、遅延が生じることがなく、信号特性値を算出するためのサンプル数を多くとることができる。

【0135】また、上記RMS [k] と、上記Min RMS [k] と、上記tone [w, k] と、上記ZC [w, k] と、上記Rxxとを、図13に示すように、例えばバックプロパゲーションタイプのニューラルネットワークの入力として用いて、雑音区間推定を行ってもよい。

【0136】上記ニューラルネットワークにおいて、上記RMS [k]、上記Min RMS [k]、上記tone [w, k]、上記ZC [w, k]、上記Rxxの各値が入力層の各端子に入力される。

【0137】入力層の各端子に入力された各値は、中間層に出力されるが、この際にシナプス荷重いわゆる重みが付けられる。すなわち、重みが乗ぜられる。

【0138】また、中間層では、それぞれに重みが付けられた各値と、バイアス51からバイアス値が入力され、所定の処理が行われた後、処理結果が出力される。

この処理結果には重みが付けられる。

【0139】出力層では、中間層から出力される重みが付けられた処理結果にバイアス52からバイアス値が入力され、所定の処理が行われた後、雑音区間推定の結果が出力される。

【0140】なお、バイアス51、52から出力される各バイアス値及び各出力に付けられる重みは、いわゆる望ましい変換を実現させるために適応的に決定される。従って、処理されるデータが多ければ多いほど確からしさが向上する。すなわち、処理が行われれば行われる程、音声と雑音との分類において、より入力音声信号に即した推定雑音レベル及びスペクトルが定まり、正確なHn値の算出を行うことができるようになる。

【0141】

【発明の効果】以上説明したように、本発明の音声信号の雑音低減方法によれば、入力される音声信号のピッチ強度に応じて雑音低減に用いるフィルタの特性を制御し、上記入力される音声信号の所定の周波数帯域、例えば高域側や低域側における雑音抑圧量を小さくすることで、上記雑音抑圧量に基づいて処理された音声信号を音声符号化しても聴覚上不自然な音声となる虞がなくなる、すなわち音質が向上する。

【図面の簡単な説明】

【図1】本発明の音声信号の雑音低減方法を適用した雑音低減装置の要部を示すブロック図である。

【図2】上記雑音低減装置のフレーム化処理部におけるフレーム化処理を説明する図である。

【図3】上記雑音低減装置の信号特性計算部におけるピッチ検出処理を説明する図である。

【図4】上記雑音低減装置におけるエネルギー[k]及び減衰エネルギーdecay[k]の具体例を示す図である。

【図5】上記雑音低減装置におけるRMS値RMS[k]、推定雑音レベル値Min RMS[k]及び最大RMS値Max RMS[k]の具体例を示す図である。

【図6】上記雑音低減装置におけるdB表示の相対エネルギーdBrel[k]、最大SN比Max SNR[k]、及び雑音判別の閾値の1つであるdBthreshrel[k]の具体例を示す図である。

【図7】上記雑音低減装置における最大SN比Max SNR[k]に対して定義される関数としてのNR_level[k]を示すグラフである。

【図8】上記雑音低減装置のadj値計算部にて得られるadj3[w, k]と周波数との関係を示すグラフである。

【図9】上記雑音低減装置における入力信号スペクトルの周波数領域の分布を示す値を求める方法を説明する図である。

【図10】上記雑音低減装置のCE値及びNR値計算部にて得られるNR[w, k]と、Hn値計算部にて得ら

25

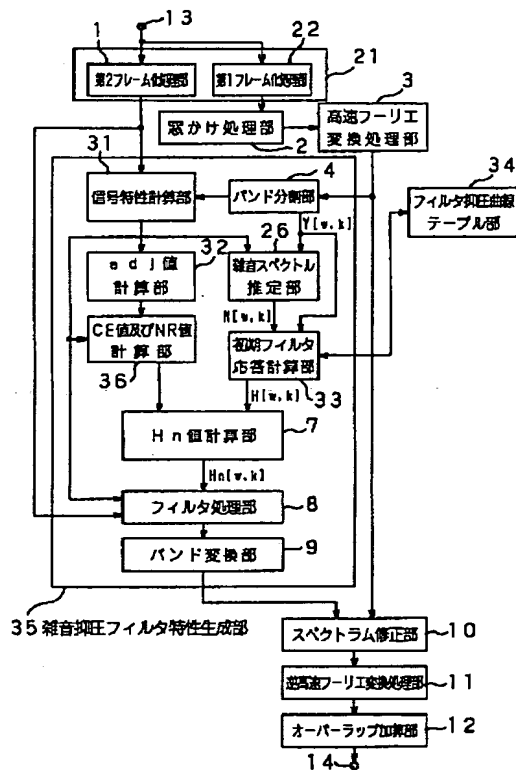
れる最大抑圧量との関係を表すグラフである。

【図11】上記雑音低減装置の出力を用いる例としての符号励振線形予測符号化のアルゴリズムによる符号化装置の要部を示すブロック図である。

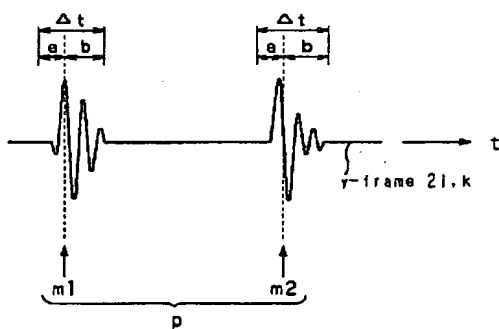
【図12】上記符号化装置により符号化された音声信号を復号化するための復号化装置の要部を示すブロック図である。

【図13】本発明の音声信号の雑音低減方法において、雑音区間推定を行う実施例を示す図である。

【図1】



【図3】

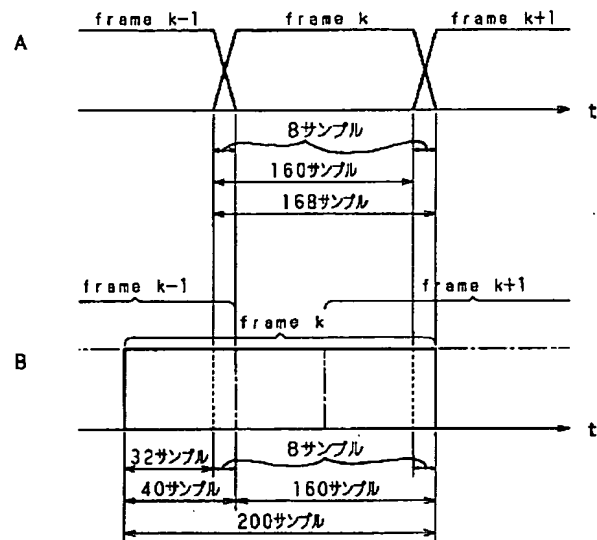


26

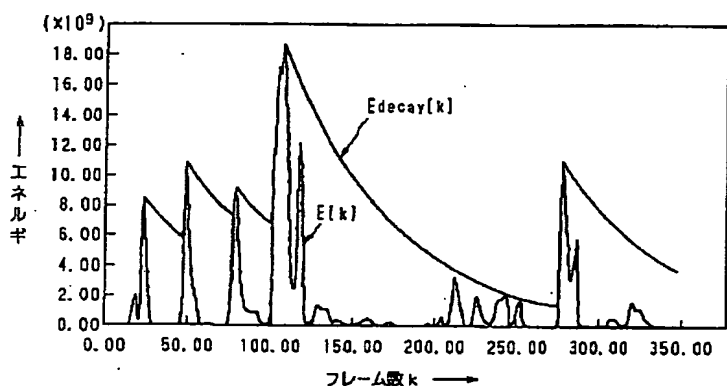
【符号の説明】

- 1 第2フレーム化処理部
- 7 H_n 値計算部
- 21 フレーム化処理部
- 22 第1フレーム化処理部
- 31 信号特性計算部
- 32 adj 値計算部
- 36 CE 値及び NR 値計算部

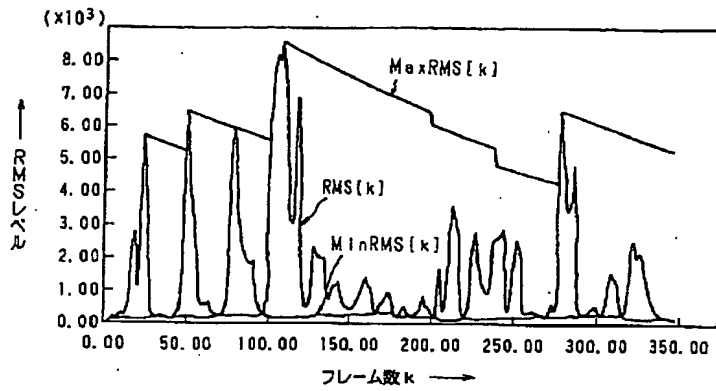
【図2】



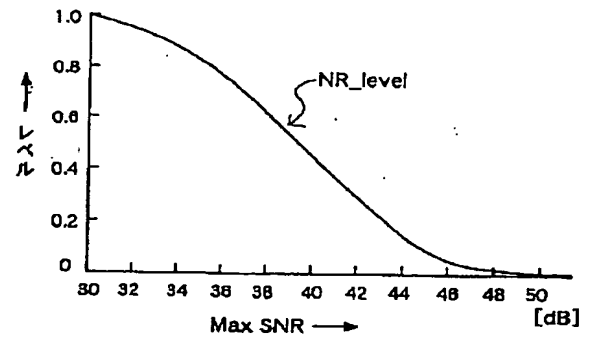
【図4】



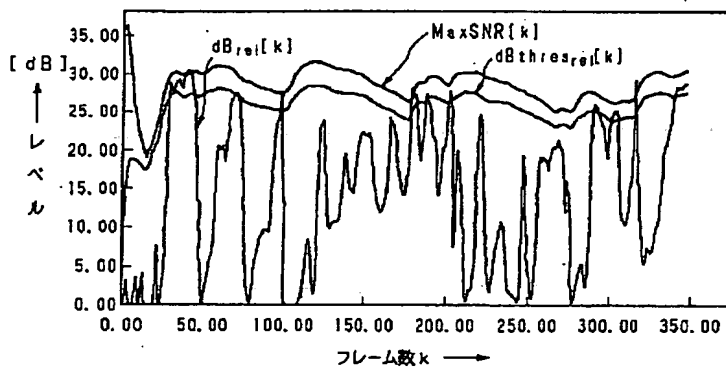
【図5】



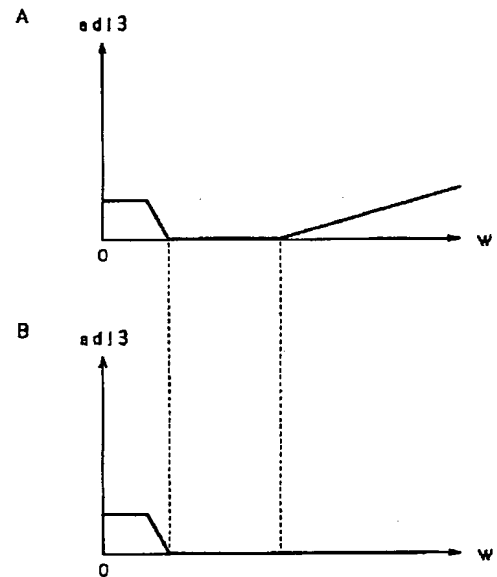
【図7】



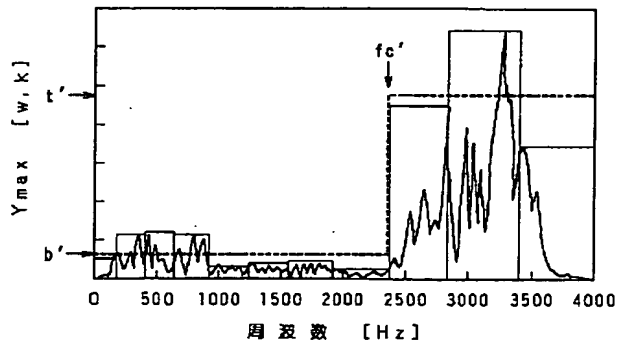
【図6】



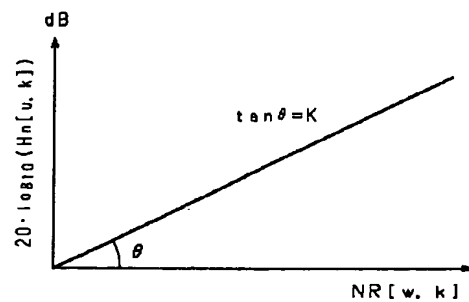
【図8】



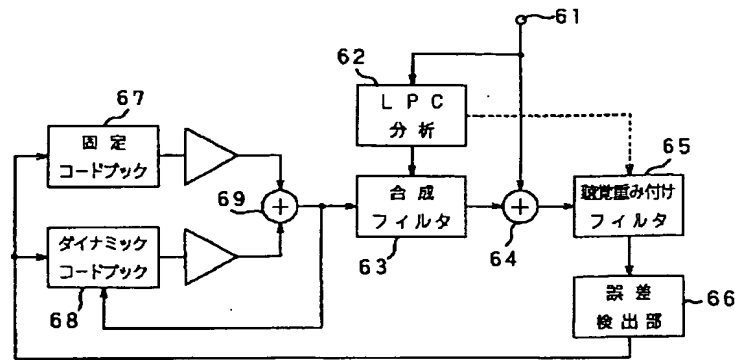
【図9】



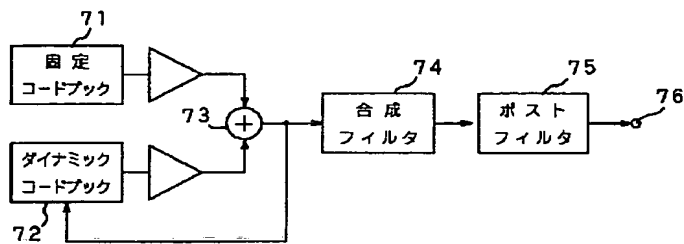
【図10】



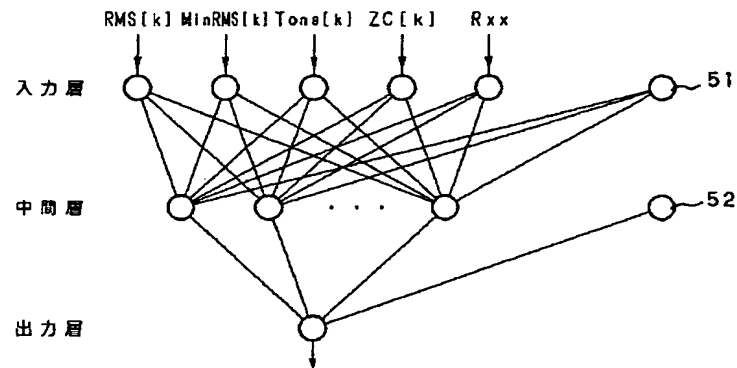
【図 11】



【図 12】



【図 13】



フロントページの続き

(51) Int. Cl.⁶

G 1 0 L 9/10

H 0 3 H 17/02

21/00

識別記号

3 0 1

6 0 1

片内整理番号

8842-5 J

8842-5 J

F I

G 1 0 L 9/10

H 0 3 H 17/02

21/00

技術表示箇所

3 0 1 C

6 0 1 H